JOURNAL OF MULTIVARIATE ANALYSIS 54, 113–125 (1995)

Nonnegative Minimum Biased Quadratic Estimation in the Linear Regression Models

S. GNOT*

Pedagogical University, Zielona Góra, Poland

G. TRENKLER†

University of Dortmund, Germany

AND

R. ZMYŚLONY*

Technical University, Zielona Góra, Poland

In the paper the problem of nonnegative estimation of $\beta' H\beta + h\sigma^2$ in the linear model $E(y) = X\beta$, $\operatorname{Var}(y) = \sigma^2 I$ is discussed. Here H is a nonnegative definite matrix while h is a nonnegative scalar. An iterative procedure for the nonnegative minimum biased quadratic estimator is described. Moreover, in the case that H and $X'X$ commute, an explicit formula for this estimator is given. Admissibility of the estimator is proved. The results are applied to nonnegative estimation of the total mean squared error of a linear biased estimator.

1. INTRODUCTION

Let us consider a linear model $M\{y, X\beta, \sigma^2 I\}$, where y is an $n \times 1$ normally distributed vector of observations, with the expectation vector $E(y) = X\beta$ and the covariance matrix $\operatorname{Cov}(y) = \sigma^2 I$. The nonstochastic $n \times p$ matrix X is known and of rank $p - q$, $0 \leqslant q < p < n$, β is a $p \times 1$ vector of unknown parameters, while $\sigma^2 > 0$ is the unknown variance of the disturbances.

Received February 2, 1993; revised October, 1994.

AMS 1980 subject classification: 62J05, 62H12.

Key words and phrases: linear model, nonnegative minimum biased estimator, admissibility, total mean squared error, one-way classification model.

* Partially supported by the Committee of Scientific Research PB 678-2-91.

† Support by Deutsche Forschungsgemeinschaft, Grant TR 253/1-2 is gratefully acknowledged.

0047-259X/95 $12.00

In the paper we consider the problem of nonnegative estimation of $\beta' H\beta + h\sigma^2$. Here H is a nonnegative definite matrix, while h is a nonnegative scalar. The problem arises, for instance, if we want to estimate accuracy of linear estimators Ly of β by means of the total mean squared error (TMSE), which has the structure $\beta' H\beta + h\sigma^2$ and it is given by

$$\begin{aligned}\mathrm{TMSE}(Ly) &= E[(Ly-\beta)'(Ly-\beta)] \\ &= \beta'(X'L'-I)(X'L'-I)'\beta + \sigma^2 \operatorname{tr}\{LL'\}. \end{aligned} \tag{1}$$

The estimators of TMSE are used for comparison of linear estimators as well as for variable selection in linear regression models (see Sen and Srivastava, 1990, p. 238; Berger and Robert, 1990).

It is well known that the ordinary least squares estimator $X\hat{\beta}$ with

$$\hat{\beta} = (X'X)^+ X'y \tag{2}$$

and the statistic

$$\hat{\sigma}^2 = y'My/(n-p+q) \tag{3}$$

with $M = I - XX^+$, are the best linear unbiased estimators of $X\beta$ and σ^2, respectively, in $M\{y, X\beta, \sigma^2 I\}$. However for estimation of the nonnegative quadratic estimable function $\beta' H\beta + h\sigma^2$ $(PHP = H,\ P = X'(XX')^+ X,\ H \geqslant 0,\ h \geqslant 0)$ the naive estimator

$$\hat{\beta}' H\hat{\beta} + h\hat{\sigma}^2 = y'X(X'X)^+ H(X'X)^+ X'y + hy'My/(n-p+q) \tag{4}$$

has bias $\sigma^2 \operatorname{tr} H(X'X)^+$. On the other hand the estimator

$$\hat{\beta}' H\hat{\beta} + \hat{\sigma}^2[h - \operatorname{tr} H(X'X)^+] \tag{5}$$

is unbiased for $\beta' H\beta + h\sigma^2$ and has uniform minimum variance as a function of the minimal sufficient and complete statistics $P\hat{\beta}$ and $\hat{\sigma}^2$. It is also nonnegative by construction if $h \geqslant \operatorname{tr} H(X'X)^+$.

Unfortunately, the estimator (5) is unacceptable in practice if $h < \operatorname{tr} H(X'X)^+$ since its values can be negative. This happens if we wish to estimate the weighted squared length $\beta' H\beta$. It can also occur if we estimate TMSE given by (1). Thus the problem arises (cf. McDonald and Galarneau, 1975; Brook and Moore, 1980; Trenkler, 1981, pp. 83, 84; and Rukhin, 1987) how to estimate $\beta' H\beta + h\sigma^2$ where H is nonnegative definite and $h \geqslant 0$, by quadratic forms $y'Ay$ with nonnegative definite A, in the case $h < \operatorname{tr}\{H(X'X)^+\}$.

2. Minimum Biased Quadratic Estimator for $\beta' H\beta + h\sigma 2$

For a given integer k let R^k, L_k and L_k^+ denote the k-dimensional Euclidean vector space, the linear space of symmetric $k \times k$ matrices and the convex cone of nonnegative definite $k \times k$ matrices, respectively.

For the function $\beta' H\beta + h\sigma^2$ which is not nonnegative estimable in the model $M\{y, X\beta, \sigma^2 I\}$ i.e. for which $h < \mathrm{tr}\{H(X'X)^+\}$ we can use Hartung's (1981) approach to find the minimum biased estimator in the class of all quadratic forms $y'Ay$ with $A \in L_n^+$, or by sufficiency of $P\hat{\beta}$ and $\hat{\sigma}^2$, in the class of the estimators being represented by $\{\hat{\beta}' C\hat{\beta} + c\hat{\sigma}^2 \mid C \in L_p^+, c \geqslant 0\}$.

Definition 2.1. We say that $\hat{\beta}' C_H \hat{\beta} + c_H \hat{\sigma}^2$, $C_H \in L_p^+$, $c_H \geqslant 0$, is a nonnegative minimum biased estimator for $\beta' H\beta + h\sigma^2$, $h < \mathrm{tr}\, H(X'X)^+$, in the model $M\{y, X\beta, \sigma^2 I\}$ if the pair (C_H, c_H) solves the following problem

$$\min_{C \in L_p^+,\, c \geqslant 0} \mathrm{tr}\{(H - C)^2\} + [\mathrm{tr}\{C(X'X)^+\} + c - h]^2. \tag{6}$$

For a given $\bar{C}$ such that $\mathrm{tr}\{\bar{C}(X'X)^+\} < h$ take $C = H - \alpha(H - \bar{C}) = (1-\alpha) H + \alpha\bar{C}$, where $0 < \alpha < 1$ is such that $\mathrm{tr}\, C(X'X)^+ = h$. In view of $\mathrm{tr}\{H(X'X)^+\} > h$ such an α exists. Note that $\mathrm{tr}(H - C)^2 = \alpha^2 \,\mathrm{tr}\{(H - \bar{C})^2\} < \mathrm{tr}\{(H - \bar{C})^2\}$ while $[\mathrm{tr}\{C(X'X)^+\} - h]^2 = 0$. It follows that

$$\begin{aligned} &\mathrm{tr}\{(H - C)^2\} + [\mathrm{tr}\{C(X'X)^+\} - h]^2 \\ &\qquad \leqslant \mathrm{tr}\{(H - \bar{C})^2\} + [\mathrm{tr}\{C(X'X)^+\} + c - h]^2. \end{aligned} \tag{7}$$

for every nonnegative c. Thus if $\mathrm{tr}\{H(X'X)^+\} > h$, in search for the solution of (6) we can restrict our considerations to $\hat{\beta}' C\hat{\beta}$. Hence the problem reduces to

$$\min_{C \in L_p^+} \rho_H(C), \tag{8}$$

where

$$\rho_H(C) = \mathrm{tr}\{(H - C)^2\} + [\mathrm{tr}\{C(X'X)^+ - h]^2\}. \tag{9}$$

A straightforward modification of the convex program leads to definition of the Lagrange function

$$L_H(C, B) = \rho_H(C) - \mathrm{tr}(CB), \tag{10}$$

where B is a nonnegative "Lagrange multiplier." Following Hartung (1981, Theorem 3.1) there exists a nonnegative B_H such that for a solution C_H of (9) we have $\operatorname{tr}\{C_H B_H\} = 0$ and

$$\rho_H(C_H) = L_H(C_H, B_H) = \min_{C \in L_p} L_H(C, B_H). \tag{11}$$

Further (C_H, B_H) is a saddle point of $L_H(C, B_H)$ on $L_p \times L_p^+$.

It can be seen that the gradient of $L_H(C, B_H)$ with respect to C vanishes if and only if

$$B_H = 2C + 2[\operatorname{tr}\{C(X'X)^+\} - h](X'X)^+ - 2H. \tag{12}$$

Generally it is not easy to find an explicit solution of (12). However it can readily be done when H and $X'X$ commute. Then there exists an orthogonal matrix E such that

$$EHE' = \operatorname{diag}(\gamma_1, ..., \gamma_p)$$

$$E(X'X)^+ E' = \operatorname{diag}(\lambda_1, ..., \lambda_p).$$

Let $(X'X)^+ = \sum_{i=1}^p \lambda_i E_i$ and $H = \sum_{i=1}^p \gamma_i E_i$ be the spectral decomposition of $(X'X)^+$ and H, respectively. Since $\operatorname{rank}(X) = p - q$, we assume that $\lambda_1 = \cdots = \lambda_q = 0$, while the remaining eigenvalues $\lambda_{q+1}, ..., \lambda_p$ and $\gamma_{q+1}, ..., \gamma_p$ are ordered such that $\gamma_{q+1}/\lambda_{q+1} \geqslant \gamma_{q+2}/\lambda_{q+2} \geqslant \cdots \geqslant \gamma_p/\lambda_p$. For a given integer k, $q < k \leqslant p$, set

$$g(k) = \sum_{l=q+1}^{k} \gamma_l \lambda_l, \qquad f(k) = \sum_{l=q+1}^{k} \lambda_l^2,$$

$$t(k) = \frac{g(k) - h}{1 + f(k)}. \tag{13}$$

$$\delta_i(k) = \gamma_i - t(k)\,\lambda_i = [\gamma_i/\lambda_i - t(k)]\,\lambda_i, \qquad i = q+1, ..., p.$$

Note that $t(k) = \sum_{l=q+1}^p \delta_l(k)\,\lambda_l - h$.

Let k_* be the maximal number k among $q+1, ..., p$ such that $\delta_k(k) > 0$. Since $\delta_{q+1}(q+1) = (\gamma_{q+1} + h\lambda_{q+1})/(1 + \lambda_{q+1}^2) > 0$, the integer k_* is uniquely defined.

LEMMA 2.1. *$\delta_i(k_*) > 0$ for $q+1 \leqslant i \leqslant k_*$ and $\delta_i(k_*) \leqslant 0$ for $i > k_*$.*

Proof. Since $\gamma_i/\lambda_i \geqslant \gamma_{k_*}\lambda_{k_*}$ for $q+1 \leqslant i \leqslant k_*$, we obtain

$$\delta_i(k_*) = \left[\frac{\gamma_i}{\lambda_i} - t(k_*)\right]\lambda_i \geqslant \left[\frac{\gamma_{k_*}}{\lambda_{k_*}} - t(k_*)\right]\lambda_i, \qquad i = q+1, ..., k_*.$$

Note that all $\delta_i(k_*)$ given above are positive because

$$\delta_{k_*}(k_*) = \left[\frac{\gamma_{k_*}}{\lambda_{k_*}} - t(k_*)\right] \lambda_{k_*}.$$

is positive. Suppose that $\delta_{k_*+1}(k_*) > 0$. Then

$$\frac{\gamma_{k_*+1}}{\lambda_{k_*+1}} > \frac{g(k_*) - h}{1 + f(k_*)},$$

$$\gamma_{k_*+1}[1 + f(k_*)] > \lambda_{k_*+1}[g(k_*) - h],$$

$$\gamma_{k_*+1}[1 + f(k_*)] + \gamma_{k_*+1}\lambda_{k_*+1}^2 > \lambda_{k_*+1}[g(k_*) - h] + \gamma_{k_*+1}\lambda_{k_*+1}^2,$$

$$\gamma_{k_*+1}[1 + f(k_* + 1)] > \lambda_{k_*+1}[g(k_* + 1) - h],$$

$$\frac{\gamma_{k_*+1}}{\lambda_{k_*+1}} > \frac{g(k_* + 1) - h}{1 + f(k_* + 1)},$$

and consequently $\delta_{k_*+1}(k_* + 1) > 0$, in contrast to the definition of k_*. Thus $\delta_{k_*+1}(k_*) \leqslant 0$, and since γ_i/λ_i is a nonincreasing sequence we have $\delta_i(k_*) \leqslant 0$ for $i \geqslant k_*$.

THEOREM 2.1. *If $h < \operatorname{tr} H(X'X)^+$ and $HX'X = X'XH$ then the non-negative minimum biased quadratic estimator of $\beta' H\beta + h\sigma^2$ in the model $M\{y, X\beta, \sigma^2 I\}$, $\operatorname{rank}(X) = p - q$, has the form $\hat{\beta}' C_H \hat{\beta}$ with*

$$C_H = \sum_{i=q+1}^{k_*} [\gamma_i - \lambda_i t(k_*)]\, E_i, \tag{14}$$

where $(X'X)^+ = \sum_{i=1}^{p} \lambda_i E_i$ and $H = \sum_{i=1}^{p} \gamma_i E_i$ are the spectral decomposition of $(X'X)^+$ and H, respectively, $\hat{\beta} = (X'X)^+ X'y$, while the eigenvalues $\lambda_{q+1}, ..., \lambda_p$ and $\gamma_{q+1}, ..., \gamma_p$ are ordered such that $\gamma_{q+1}/\lambda_{q+1} \geqslant \gamma_{q+2}/\lambda_{q+2} \geqslant \cdots \geqslant \gamma_p/\lambda_p$.

Proof. First note that the nonnegativity of C_H follows directly from Lemma 2.1. By noting that $\operatorname{tr} C_H(X'X)^+ - h = g(k_*) - f(k_*) - h = t(k_*)$, from (12) we find that

$$B_H = 2 \sum_{i=k_*+1}^{p} [\lambda_i t(k_*) - \gamma_i]\, E_i.$$

It follows that B_H is also nonnegative definite, and the condition $\operatorname{tr}(C_H B_H) = 0$ is clearly satisfied.

Remark 2.1. Let C be any nonnegative definite $p \times p$-matrix and let $\mathscr{E} = \mathrm{sp}\{E_i, i = 1, ..., p\}$. Following Olsen *et al.* (1976) denote by Π the orthogonal projection on $\mathscr{E}$ with respect to inner product $\langle \cdot, \cdot \rangle$ in L_p given by $\langle A, B \rangle = \mathrm{tr}\{AB\}$. Then $C_0 = \Pi(C) = \sum_{i=1}^{p} (\mathrm{tr}\{CE_i\}) E_i$ is also nonnegative definite, and $C = C_0 + C_1$, where $C_1 = C - C_1 \perp \mathscr{E}$. Since $H \in \mathscr{E}$, we have $\langle H, C_1 \rangle = 0$ and if $C_1 \neq 0$, then

$$\begin{aligned} \mathrm{tr}\{(H-C)^2\} &= \langle H - C, H - C \rangle \\ &= \langle H - C_0, H - C_0 \rangle + \langle C_1, C_1 \rangle > \mathrm{tr}\{(H - C_0)^2\}. \end{aligned}$$

Moreover we have $\mathrm{tr}\{C(X'X)^+\} = \mathrm{tr}\{C_0(X'X)^+\}$. It follows that in looking for the solution of (8) we can restrict our considerations to $C \in \mathscr{E}$.

Remark 2.2. If $\mathrm{rank}(H) = \mathrm{rank}(X)$ and $k_* = p$, i.e., $\gamma_p > t(p)\,\lambda_p$, where

$$t(p) = \frac{\mathrm{tr}\{H(X'X)^+\} - h}{1 + \mathrm{tr}\{(X'X)^+ (X'X)^+\}},$$

then C_H given by (14) has the following form

$$C_h = H - t(p)(X'X)^+. \tag{15}$$

In such a case the nonnegative minimum biased quadratic estimator for $\beta' H \beta + h\sigma^2$ has the form

$$\hat{\beta}' H \hat{\beta} - t(p)\, \hat{\beta}'(X'X)^+ \hat{\beta}. \tag{16}$$

The expectation of the estimator is $\beta' H \beta - t(p)\, \beta'(X'X)^+ \beta + [t(p) + h]\, \sigma^2$.

The assumption that H and $X'X$ commute is restrictive, but as Section 4 shows, it is fulfilled in many practical situations. We describe now a procedure that leads to the solution of (8) without a commutativity assumption. In the procedure we use the trivial fact that since C_H and B_H are assumed to be nonnegative definite, $\mathrm{tr}(C_H B_H) = 0$ if and only if $C_H B_H = 0$.

For an arbitrary symmetric matrix A with the spectral decomposition $A = \sum_i a_i P_i$ let

$$A_+ = \sum_i a_i^* P_i, \qquad A_- = -\sum_i a_{*i} P_i$$

with $a_i^* = \max\{0, a_i\}$ and $a_{*i} = \min\{0, a_i\}$, being the positive and the negative part of A, respectively. The following corollary follows directly from (12).

COROLLARY 2.1. *If* $h < \operatorname{tr}\{H(X'X)^+\}$ *then the nonnegative minimum biased quadratic estimator of* $\beta' H\beta + h\sigma^2$ *in the model* $M\{y, X\beta, \sigma^2 I\}$, $\operatorname{rank}(X) = p - q$, *has the form* $\hat{\beta}' C_H \hat{\beta}$ *with*

$$C_H = [H - \lambda(X'X)^+]_+, \tag{17}$$

and $\lambda = \operatorname{tr}\{C_H(X'X)^+\} - h$. *The matrix* B_H *of Lagrange multipliers is given by* $B_H = 2[H - \lambda(X'X)^+]_-$ *and an implicit formula for* λ *is*

$$\lambda = [\operatorname{tr}\{H(X'X)^+\} - h]/d + [\operatorname{tr}\{B_H(X'X)^+\} - h]/2d, \tag{18}$$

where $d = 1 + \operatorname{tr}\{(X'X)^+ (X'X)^+\}$.

We describe now a procedure leading to the solution of (18) if $h < \operatorname{tr}\{H(X'X)^+\}$. For a fixed real number δ put $C(\delta) = H - \delta(X'X)^+$.

LEMMA 2.2. Tr $C_-(\delta)(X'X)^+$ *is a nondecreasing function of* δ.

Proof. Let $\delta_2 > \delta_1$. Since $C_+(\delta) - C_-(\delta) = H - \delta(X'X)^+$ we have

$$C_+(\delta_2) - C_+(\delta_1) - [C_-(\delta_2) - C_-(\delta_1)] = -(\delta_2 - \delta_1)(X'X)^+. \tag{19}$$

On the other hand $C_+(\delta)\, C_-(\delta) = 0$ implies

$$\{C_+(\delta_1) + [C_+(\delta_2) - C_+(\delta_1)]\}\{C_-(\delta_1) + [C_-(\delta_2) - C_-(\delta_1)]\} = 0.$$

It follows that

$$\operatorname{tr}\{C_+(\delta_1)\, C_-(\delta_2)\} + \operatorname{tr}\{C_+(\delta_2)\, C_-(\delta_1) + \operatorname{tr}\{[C_-(\delta_2) - C_-(\delta_1)]^2\}$$
$$- (\delta_2 - \delta_1) \operatorname{tr}\{[C_-(\delta_2) - C_-(\delta_1)](X'X)^+\} = 0.$$

Since the first three traces are all nonnegative and $\delta_2 - \delta_1 > 0$, we find that

$$\operatorname{tr}\{[C_-(\delta_2) - C_-(\delta_1)](X'X)^+\} \geqslant 0, \tag{20}$$

which proves the lemma.

Let us define a sequence $\{\delta_n\}$ as follows:

$$\delta_0 = [\operatorname{tr}\{H(X'X)^+\} - h]/d,$$
$$\delta_n = [\operatorname{tr}\{C_-(\delta_{n-1})(X'X)^+\}]/d + \delta_0, \qquad n = 1, 2, ... \tag{21}$$

Since $h < \operatorname{tr}\{H(X'X)^+\}$, we have $\delta_1 > \delta_0$, and by induction from Lemma 2.2 it follows that $\{\delta_n\}$ is nondecreasing sequence. Besides taking into account that $B_H = 2C_-(\lambda)$, from (18) we have $\delta_0 \leqslant \lambda$ and

$$\lambda - \delta_n = \operatorname{tr}\{C_-(\lambda)(X'X)^+\}/d - \operatorname{tr}\{C_-(\delta_{n-1})(X'X)^+\}/d \geqslant 0.$$

This implies that $\{\delta_n\}$ is bounded from above by λ and by Lemma 2.2 it follows that $\{\delta_n\}$ converges.

THEOREM 2.2. *Let* $\lim_{n\to\infty} \delta_n = \delta$, *where* δ_n *is given by* (21). *Then* $C_H = C_+(\delta)$.

Proof. Multiplying both sides of $C_+(\delta_n) - C_-(\delta_n) = H - \delta_n(X'X)^+$ by $(X'X)^+$ and taking traces we find that

$$\operatorname{tr}\{C_+(\delta_n)(X'X)^+\} = -\delta_n d_1 + \operatorname{tr}\{H(X'X)^+\} + \operatorname{tr}\{C_-(\delta_n)(X'X)^+\},$$

where $d_1 = \operatorname{tr}\{(X'X)^+ (X'X)^+\}$. Using (21) we get

$$\begin{aligned}
&\operatorname{tr}\{C_+(\delta_n)(X'X)^+\}\\
&\qquad - \operatorname{tr}\{C_-(\delta_n)(X'X)^+\}\, d_1/d\\
&\qquad + \operatorname{tr}\{H(X'X)^+\} - \operatorname{tr}\{H(X'X)^+\}\, d_1/d + \operatorname{tr}\{C_-(\delta_n)(X'X)^+\}\\
&\quad = \operatorname{tr}\{C_-(\delta_{n-1})(X'X)^+\}\, d_1/d + \operatorname{tr}\{H(X'X)^+\}/d\\
&\qquad + \operatorname{tr}\{C_-(\delta_n)(X'X)^+\}\\
&\quad = [\operatorname{tr}\{C_-(\delta_n)(X'X)^+\} - \operatorname{tr}\{C_-(\delta_{n-1})(X'X)^+\}]\, d_1/d\\
&\qquad + \operatorname{tr}\{H(X'X)^+\}/d + \operatorname{tr}\{C(\delta_n)(X'X)^+\}/d.
\end{aligned}$$

Since $\operatorname{tr}\{C_-(\delta_n)(X'X)^+\} - \operatorname{tr}\{C_-(\delta_{n-1})(X'X)^+\}$ tends to zero if n tends to infinity, we conclude from (21) that

$$\operatorname{tr}\{C_+(\delta)(X'X)^+\} = \delta_0 + \operatorname{tr}\{C_-(\delta)(X'X)^+\}/d = \delta.$$

This proves Theorem 2.2 in view of Corollary 2.1.

3. ADMISSIBILITY OF NONNEGATIVE MINIMUM BIASED QUADRATIC ESTIMATORS

One can say that a small bias may not be a good reason for using a procedure if its variance is too large. For example in the problem of estimation of variance components by invariant quadratic forms the minimum biased procedure can produce inadmissible estimators if the mean squared error (MSE) risk function is taken into account (cf. Gnot and Srzednicka, 1988). We should pay attention, however, that contrary to the problem of variance component invariant estimation, in the considered case the mean squared error of $\hat{\beta}'C\hat{\beta}+c\hat{\sigma}^2$ for estimation of $\beta'H\beta+h\sigma^2$ depends not only on σ^2, but also on β and has the form

$$\begin{aligned}\mathrm{MSE}(\hat{\beta}'C\hat{\beta}+c\hat{\sigma}^2)&\\ &=[\beta'(C-H)\beta+\sigma^2(\mathrm{tr}\{C(X'X)^+\}+c-h)]^2\\ &\quad+4\sigma^2\beta'C(X'X)^+C\beta+2\sigma^4\,\mathrm{tr}\{C(X'X)^+C(X'X)^+\}\\ &\quad+2\sigma^4c^2/(n-p+q).\end{aligned}\tag{22}$$

We prove now that if H and $X'X$ commute then there does not exist a nonnegative definite matrix C and a nonnegative scalar c such that $\hat{\beta}'C\hat{\beta}+c\hat{\sigma}^2$ has uniformly smaller MSE then the nonnegative minimum biased estimator.

THEOREM 3.1. *If $h<\mathrm{tr}\{H(X'X)^+\}$ and H and $X'X$ commute then the nonnegative minimum biased quadratic estimator of $\beta'H\beta+h\sigma^2$ in the model $M\{y, X\beta, \sigma^2I\}$ is admissible in the class of all nonnegative quadratic estimators with respect to the mean squared error risk function.*

Proof. Consider the minimum biased estimator given by (14). Let $(X'X)^+=\sum_{i=1}^p\lambda_iE_i$ and $H=\sum_{i=1}^p\gamma_iE_i$. Let C be a nonnegative definite $p\times p$-matrix and c a nonnegative scalar.

Suppose that $\hat{\beta}'C\hat{\beta}\neq\hat{\beta}'C_H\hat{\beta}$. If $\mathrm{MSE}(\hat{\beta}'C\hat{\beta}+c\hat{\sigma}^2)\leqslant\mathrm{MSE}(\hat{\beta}'C_H\hat{\beta})$ for each β and σ^2, then putting first $\sigma^2=0$ and next $\beta=0$, from (22) we have

$$[\beta'(C-H)\beta]^2\leqslant[\beta'(C_H-H)\beta]^2\tag{23}$$

for each β and

$$\begin{aligned}(\mathrm{tr}\{C(X'X)^+\}+c-h)^2+2\,\mathrm{tr}\{C(X'X)^+C(X'X)^+\}+\frac{2c^2}{n-p+q}&\\ \leqslant(\mathrm{tr}\{C_H(X'X)^+\}-h)^2+2\,\mathrm{tr}\{C_H(X'X)^+C_H(X'X)^+\},&\end{aligned}\tag{24}$$

with

$$\begin{aligned}\mathrm{tr}\{(C_H-H)^2\}+[\mathrm{tr}\{C_H(X'X)^+\}-h]^2&\\ <\mathrm{tr}\{(C-H)^2\}+[\mathrm{tr}\{C(X'X)^+\}+c-h]^2&\end{aligned}\tag{25}$$

as a minimum bias condition for $\hat{\beta}'C_H\hat{\beta}$. Note that (23) implies

$$\mathrm{tr}\{(C-H)^2E_i\}\leqslant\mathrm{tr}\{(C_H-H)^2E_i\}.$$

Using similar arguments as in Remark 2.1, without loss of generality we can assume that $C=\sum_{i=1}^{p} c_i E_i \in \mathscr{E}$. Then from above it follows that

$$(c_i-\gamma_i)^2 \leqslant t^2(k_*)\,\lambda_i^2, \qquad i=q+1, ..., k_*, \tag{26}$$

while from (23) and (25) we get

$$[\operatorname{tr}\{C_H(X'X)^+\}-h]^2<[\operatorname{tr}\{C(X'X)^+\}+c-h]^2. \tag{27}$$

Using now (24) and (27) we conclude that

$$\operatorname{tr}\{C(X'X)^+\,C(X'X)^+\}<\operatorname{tr}\{C_H(X'X)^+\,C_H(X'X)^+\}. \tag{28}$$

From (28) it follows that

$$\sum_{i=q+1}^{p} c_i^2\lambda_i^2<\sum_{i=q+1}^{k_*}[\gamma_i-t(k_*)\,\lambda_i]^2\,\lambda_i^2, \tag{29}$$

while (26) leads to

$$\gamma_i-t(k_*)\,\lambda_i\leqslant c_i\leqslant\gamma_i+t(k_*)\,\lambda_i, \qquad i=q+1, ..., k_*. \tag{30}$$

Since $c_i\geqslant 0$ for $i=1, ..., p$, from (30) we find that $c_i\geqslant\gamma_i-t(k_*)\,\lambda_i$ for $i=q+1, ..., k_*$, and

$$\sum_{i=q+1}^{p} c_i^2\lambda_i^2\geqslant\sum_{i=q+1}^{k_*}[\gamma_i-t(k_*)\,\lambda_i]^2\,\lambda_i^2.$$

which contradicts (29).

4. Examples

Example 4.1. Let us consider n random variables

$$y_i=\mu+\varepsilon_i, \qquad i=1, ..., n,$$

where μ is an unknown mean, common for each y_i, while $E(\varepsilon_i)=0$, $E(\varepsilon_i^2)=\sigma^2$. The matrix form of the simple model is $y=1_n\mu+\varepsilon$, $E(y)=1_n\mu$, $\operatorname{Var}(y)=\sigma^2 I_n$, $(X'X)^{-1}=1/n$, and according to Remark 2.2 the nonnegative minimum biased quadratic estimator for μ^2 is $\hat{\mu}^2=(\sum_{i=1}^{n} y_i)^2)/(1+n^2)$. The expected value of this estimator is $(n^2\mu^2+n\sigma^2)/(1+n^2)$, and its variance is $(2n^2\sigma^4+4n^3\mu^2\sigma^2)/(1+n^2)^2$. Hence $\hat{\mu}^2$ converges to μ^2 in quadratic mean. Observe that the $UMVU$ estimator for μ^2 is given by $\bar{\mu}^2=\bar{y}^2-\sum_{i=1}^{n}(y_i-\bar{y})^2/n(n-1)$ and has the variance $[2\sigma^4+4(n-1)\,\mu^2\sigma^2]/n(n-1)$. Lehmann (1983, p. 114) comments on $\bar{\mu}^2$ as follows: "... can take on negative values although the estimand is known to be non-negative. Except when $\mu = 0$ or n is small, the probability of such values is not large, but when they do occur they cause an embarrassment. This difficulty can be avoided by replacing it by zero whenever it leads to a negative value; the resulting estimator of course, will no longer be unbiased." For further discussion see Ruhkin (1987).

EXAMPLE 4.2. Let us consider the one-way classification model

$$y_{ij} = \mu_i + \varepsilon_{ij}, \qquad i = 1, ..., p, \quad j = 1, ..., r_i,$$

where $\mu = (\mu_1, ..., \mu_p)'$ is a vector of unknown means in p cells, $r = (r_1, ..., r_p)'$ is a vector for replications in cells, while $\sum_{i=1}^{p} r_i = n$ is a total number of observations. The above model can be presented in the following matrix from

$$y = \text{diag}\{1_{r_i}\}\, \mu + \varepsilon, \qquad E(y) = \text{diag}\{1_{r_i}\}\, \mu + \varepsilon, \qquad \text{Var}(y) = \sigma^2 I_n.$$

It is easy to check that

$$(X'X)^{-1} = \text{diag}\{1/r_1, ..., 1/r_p\}.$$

If the r_i's are such that $k_* = p$ then by Remark 2.2 the nonnegative minimum biased estimator for $\mu'\mu = \sum_{i=1}^{p} \mu_i^2$ is $\hat{\mu}' C_I \hat{\mu}$, where $\hat{\mu} = (\bar{y}_1, ..., \bar{y}_p)'$ with $\bar{y}_i = \sum_{j=1}^{r_i} y_{ij}/r_i$ being the mean in the i-th cell, $C_I = I - t(p)(X'X)^{-1}$, while

$$t(p) = \frac{\sum_{i=1}^{p} (1/r_i)}{1 + \sum_{i=1}^{p} (1/r_i^2)}.$$

Consequently we get

$$\hat{\mu}' C_I \hat{\mu} = \sum_{i=1}^{p} \frac{r_i - t(p)}{r_i} \bar{y}_i^2.$$

The expectation of $\hat{\mu}' C_I \hat{\mu}$ is

$$E(\hat{\mu}' C_I \hat{\mu}) = \sum_{i=1}^{p} \frac{r_i - t(p)}{r_i} \mu_i^2 + t(p)\, \sigma^2,$$

and it tends to $\mu'\mu$ if r_i tends to infinity for each i. Note that $k_* = p$ is equivalent to the condition

$$(1/r_{\min}) \sum_{i=1}^{p} \frac{1}{r_i} < 1 + \sum_{i=1}^{p} \frac{1}{r_i^2},$$

where $r_{\min}$ is the minimum of the integers $r_1, ..., r_p$. This condition my be hurt if for instance $r_i = i$ and $p > 6$. If however p is fixed and r_i is sufficient large for each i, then the condition is also valid. In the balanced case $r_i = p$ we clearly have $k_* = p$, and the nonnegative minimum biased estimator for $\mu'\mu$ becomes

$$\frac{p^2}{p^2+p} \sum_{i=1}^{p} \bar{y}_i^2.$$

EXAMPLE 4.3. Let L be a $p \times n$-matrix of constants. Then the total mean squared error of linear estimator Ly of β in the model $M\{y, X\beta, \sigma^2 I\}$ is defined as follows

$$\beta'(X'L'-I)(X'L'-I)'\beta + \sigma^2[h - \mathrm{tr}\{(LL')\}].$$

It has been proved by LaMotte (1978, Theorem 1) that if Ly is an admissible linear estimator of β, then there exists a $p \times p$-matrix D such that $L = DX'$. In particular choosing $D = (kI + X'X)^{-1}$, $k \geq 0$, we get the admissible estimator

$$\bar{\beta} = (X'X + kI)^{-1} X'y,$$

which is also known in literature as a ridge estimator of β. The TMSE of $\bar{\beta}$ has the form (cf. Trenkler, 1981, p. 120)

$$\mathrm{TMSE}(\bar{\beta}) = k^2\beta'(X'X+kI)^{-2}\beta + \sigma^2 \,\mathrm{tr}\{X'X(X'X+kI)^{-2}\} = \beta' H\beta + h\sigma^2,$$

with $H = k^2(X'X+kI)^{-2}$ and $h = \mathrm{tr}\{X'X(X'X+kI)^{-2}\}$. It follows that $\mathrm{tr}\{H(X'X)^{-1}\} \leq h$ iff $k^2 \,\mathrm{tr}\{(X'X+kI)^{-2}(X'X)^{-1}\} \leq \mathrm{tr}\{X'X(X'X+kI)^{-2}\}$. Let $(X'X)^{-1} = \sum_{i=1}^{p} \lambda_i E_i$ be the spectral decomposition of $(X'X)^{-1}$. Then the above inequality is satisfied iff

$$k^2 \sum_{i=1}^{p} \left(\frac{1}{\lambda_i} + k\right)^{-2} \lambda_i \leq \sum_{i=1}^{p} \left(\frac{1}{\lambda_i} + k\right)^{-2} \Big/ \lambda_i \tag{31}$$

which holds for sufficiently small k. For example in the balanced one-way classification model this is valid iff $k \leq p$. If (31) holds then the uniformly minimum variance estimator for $\mathrm{TMSE}(\bar{\beta})$ is nonnegative, while if (31) is hurt, the nonnegative minimum biased estimator for $\mathrm{TMSE}(\bar{\beta})$ is given by Theorem 2.1.

REFERENCES

[1] BERGER, J. O., AND ROBERT, C. (1990). Subjective hierarchical bayes estimation of a multivariate normal mean: on the frequentist interface. *Ann. Statist.* **18** 617–651.

[2] BROOK, R. J., AND MOORE, T. (1980). On the expected length of the least squares coefficient vector. *J. Econometrics* **12** 245–246.

[3] GNOT, S., AND SRZEDNICKA, J. (1988). On minimum biased quadratic estimators. *Statistics* **19** 215–222.

[4] HARTUNG, J. (1981). Nonnegative minimum biased invariant estimation in variance component models. *Ann. Statist.* **9** 278–292.

[5] LAMOTTE, L. R. (1978). Bayes linear estimators. *Technometrics* **20** 281–290.

[6] LEHMANN, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

[7] MCDONALD, G. C., AND GALARNEAU, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *J. Amer. Statist. Assoc.* **70** 407–416.

[8] OLSEN, A., SEELY, J., AND BIRKES, D. (1976). Invariant quadratic unbiased estimation for two variance components. *Ann. Statist.* **4** 878–890.

[9] RUKHIN, A. L. (1987). Quadratic estimators of quadratic functions of normal parameters. *J. Statist. Plann. Infer.* **15** 301–310.

[10] SEN, A., AND SRIVASTAVA, M. S. (1990). *Regression Analysis: Theory, Methods, and Applications.* Springer-Verlag, New York/Heidelberg.

[11] TRENKLER, G. (1981). *Biased Estimators in the Linear Regression Model.* Mathematical Systems in Economics, Vol. 58. Oelgeschlager, Gunn & Hain, Cambridge, MA.